

АВТОМАТИЗОВАНА ПОБУДОВА СЕМАНТИЧНОЇ МЕРЕЖІ НА ОСНОВІ ТЛУМАЧНОГО СЛОВНИКА

Петрасова Світлана Валентинівна, Хайрова Ніна Феліксівна

Національний технічний університет «Харківський політехнічний інститут»

Харків, Україна

svetapetrasova@gmail.com, nina_khajrova@yahoo.com

Семантична мережа сьогодні є найбільш перспективним способом формального вираження знань. Це обумовлено рядом причин. По-перше, характерною особливістю семантичних мереж є наочність представлення знань. Також їх цінність полягає в можливості врахування в базі знань, крім форм тверджень, їх семантики [1]. Семантичні мережі використовуються для розширення електронних тезаурусів; як частина лінгвістичного процесора в інформаційно-пошукових системах; в системах обробки природної мови (ПМ), експертних системах і т.д.

Але, незважаючи на досить широке використання даного способу представлення знань, їх поширення стримується як неоднозначністю представлення ПМ, так і трудомісткістю та складністю розробки. У своїй переважній більшості семантичні мережі розробляються експертами в тій чи іншій предметній області. Тоді як одним з найбільш повних джерел знань є такі універсальні засоби представлення, накопичення і передачі інформації у людському суспільстві, як тексти. Серед усіх текстових джерел тлумачні словники представляють тексти природної мови, мабуть, з найбільш концентрованим смисловим навантаженням.

У даному дослідженні пропонується метод автоматизованої побудови фрагмента семантичної мережі на базі знань тлумачного словника англійської мови. У розробленій мережі використовуються відношення класифікації, гіпонімії, меронімії, гіперонімії та відношення семантичної еквівалентності.

Для формалізації відношень семантичної еквівалентності визначається міра семантичної близькості, формально обумовлена дефініціями глосаріїв як відношення потужностей множин, утворених теоретико-множинним перетином і об'єднанням множин термінів дефініцій [2].

Для формалізації інших міжконцептуальних відношень розробленого фрагмента семантичної мережі використовуються шаблони лексичних послідовностей. Наприклад, формальним вираженням відношення класифікації може бути шаблон $NN_{MoC} - Rel - NN_{Class}$, де MoC – член класу, $Class$ – клас, Rel – лексичні ланцюжки, які виражають відношення класифікації (*meaning of, identification of, is a* та ін.).

У результаті дослідження розроблено та програмно реалізовано алгоритм автоматизованої побудови семантичної мережі, що виявляє вищенаведені відношення між термінами тлумачного словника.

Для визначення ефективності розробленого інформаційно-лінгвістичного забезпечення досліджувалася вибірка зі ста термінів словника Microsoft Computer Dictionary, 5th edition. Після обробки результатів були визначені коефіцієнти точності, повноти і помилки.

Література: 1 Маннинг Кристофер Д., Рагхаван Прабхакар, Шютце Хайнрих Введение в информационный поиск: Пер. с англ. – М.: ООО "И.Д.Вильямс", 2011. – 528с. 2 Федорченко Л.А. Метод автоматизированного построения семантической сети терминов учебной дисциплины /Л.А. Федорченко Н.Ф. Хайрова, А.И. Довнар, С.О. Булгаков. // Радиоэлектронні і комп'ютерні системи, 2011, №4 (52). С. 1-7.